**Miao Fan** [1] **DOB: Nov. 20th, 1991**

---


CONTACT INFORMATION

(1) 41-35 Elbertson St, Elmhurst, NY 11373, U.S.
(2) FIT, Tsinghua University, Beijing, 100084, P.R. China.

*Phone:* (1) +1(347)-891-8622 (U.S.)
(2) +86-13581700448 (China)
*E-mail:* fanmiao.cslt.thu@gmail.com
*WWW:* http://cslt.riit.tsinghua.edu.cn/mediawiki/index.php/Miao_Fan


RESEARCH INTERESTS

**Text Mining & Social Computing**

- *Information Extraction*
- *Social Tag Recommendation*

EDUCATION

**New York University**, New York, U.S.A. **March 2015 - March 2016**
*Joint-supervised Ph.D. Candidate in Computer Science.*
*Junior Research Scientist, Courant Institute of Mathematical Sciences.*

- Research Topic: "Large-scale Entity Relation Extraction based on Low-dimensional Representations" (Proteus Project)[2].
- Advisor: Prof. Ralph Grishman[3]

**Tsinghua University**, Beijing, P.R. China. **September 2012 - June 2017 (expected)**
*Ph.D. Candidate in Computer Science.*

- Dissertation Topic: "Large-scale Knowledge Base Completion based on Low-dimensional Representations".
- Co-advisors: Prof. Thomas Fang Zheng[4]and Prof. Qiang Zhou[5].
- GPA: **87.3**/100.

**Beijing University of Posts and Telecommunications September 2008 - June 2012**
*B.Eng. in Software Engineering.*

- Dissertation Topic: "Chinese Natural Question Generation Based on Wiki Knowledge"[6], **Best B.Eng. Dissertation Award**.
- Advisor: Prof. Guoshi Wu[7].
- GPA: **95.4**/100 (Rank **1**/97).

MY MONOGRAPHS

**Miao Fan**, Chao Li; *DIY Machine Learning Systems for Kaggle Competitions with Python Programming*, Printing, Tsinghua University Press.

**Miao Fan**, *Distributed Machine Learning and Big Data Analysis with PySpark*, Drafting, Tsinghua University Press.

PATENTS

**Miao Fan**, Doo Soon Kim. Detecting Table Region in PDF Documents using Distant Supervision[8]. U.S. Patent submitted.

**Miao Fan**, Qiang Zhou. The Approach to Social Tag Generation and Recommendation based on Monolingual Word Alignments. P.R.China Patent submitted.

---

[1] http://scholar.google.com/citations?user=aPlHReAAAAAJ&hl=en
[2] http://nlp.cs.nyu.edu/index.shtml
[3] http://scholar.google.com/citations?user=blwKAkUAAAAJ
[4] http://scholar.google.com/citations?user=H3MX_8IAAAAJ&hl=en
[5] http://cslt.riit.tsinghua.edu.cn/~qzhou/eng/index.htm
[6] http://cslt.riit.tsinghua.edu.cn/mediawiki/images/a/a9/B.ENG_Dissertation_-Miao_Fan-.pdf
[7] http://baike.baidu.com/view/9021485.htm
[8] http://arxiv.org/abs/1506.08891

PUBLICATIONS

**Miao Fan**, Qiang Zhou, Thomas Fang Zheng, Ralph Grishman. Distributed Representation Learning for Knowledge Bases with Entity Descriptions. *Pattern Recognition Letters*[9] (**IF:1.55**), Elsevier. SCIE Index. Full length article submitted.

**Miao Fan**, Qiang Zhou, Andrew Abel, Thomas Fang Zheng, Ralph Grishman. Probabilistic Belief Embedding for Large-scale Knowledge Population. *Cognitive Computation*[10] (**IF: 1.44**), Springer. SCIE Index. Full length article submitted.

**Miao Fan**, Qiang Zhou, Thomas Fang Zheng. Learning Embedding Representations for Imperfect Knowledge Repository.[11] 2016 IEEE/WIC/ACM International Conference on Web Intelligence (*WI'16*). Full research paper submitted.

**Miao Fan**, Qiang Zhou, Thomas Fang Zheng. Distant Supervision for Entity Linking.[12] The 29th Pacific Asia Conference on Language, Information and Computation (*PACLIC'15*), pp. 79-86. Full paper, oral presentation.

**Miao Fan**, Kai Cao, Yifan He, Ralph Grishman. Jointly Embedding Relations and Mentions for Knowledge Population.[13] The 10th Recent Advances in Natural Language Processing (*RANLP'15*), pp. 186-191. Poster paper.

**Miao Fan**, Qiang Zhou, Thomas Fang Zheng, Ralph Grishman. Large Margin Nearest Neighbor Embedding for Knowledge Representation.[14] The 2015 IEEE/WIC/ACM Web Intelligence Conference (*WI'15*), pp. 53-59, 6-9 December 2015, Singapore. Full paper, oral presentation.

**Miao Fan**, Deli Zhao, Qiang Zhou, Zhiyuan Liu, Thomas Fang Zheng, Edward Y. Chang. Distant Supervision for Relation Extraction with Matrix Completion.[15] The 52th Annual Meeting of the Association for Computational Linguistics (*ACL'14*), pp. 839-849. Full paper, oral presentation.

**Miao Fan**, Qiang Zhou, Emily Chang, Thomas Fang Zheng. Transition-based Knowledge Graph Embedding with Relational Mapping Properties.[16] The 28th Pacific Asia Conference on Language, Information and Computing (*PACLIC'14*), pp. 328-337. Full paper, oral presentation.

**Miao Fan**, Qiang Zhou, Thomas Fang Zheng. Mining the Personal Interests of Microbloggers via Exploiting Wikipedia Knowledge.[17] 15th International Conference on Intelligent Text Processing and Computational Linguistics (*CICLing'14*), pp. 188-200. Full paper, poster presentation.

**Miao Fan**, Qiang Zhou, Thomas Fang Zheng. Content-based Semantic Tag Ranking for Recommendation.[18] The 2012 IEEE/WIC/ACM International Conference on Web Intelligence (*WI'12*), pp. 292-296. Short paper, oral presentation.

**Miao Fan**, Yingnan Xiao, Qiang Zhou. Bringing the Associative Ability to Social Tag Recommendation.[19] *ACL'12 Workshop* on Graph-based Methods for Natural Language Processing, pp. 44-54. Workshop paper, oral presentation.

**Miao Fan**, Guoshi Wu. Aspect Opinion Mining on Customer Reviews.[20] Proceedings of the 2011 International Conference on Informatics, Cybernetics, and Computer Engineering (*ICCE'11*) November 19-20, 2011, Melbourne, Australia. Advances in Intelligent and Soft Computing Volume 112, 2012, pp 27-33.

[9] http://www.journals.elsevier.com/pattern-recognition-letters/
[10] http://link.springer.com/journal/12559
[11] http://arxiv.org/pdf/1503.08155v1.pdf
[12] http://aclweb.org/anthology/Y15-1010
[13] http://arxiv.org/pdf/1504.01683v1.pdf
[14] http://arxiv.org/pdf/1504.01684v1.pdf
[15] http://arxiv.org/abs/1411.4455
[16] http://pan.baidu.com/s/1pJqLPIb
[17] http://pan.baidu.com/s/1bntmkpx
[18] http://pan.baidu.com/s/1kT41lyf
[19] http://pan.baidu.com/s/1o6wn7lg
[20] http://pan.baidu.com/s/1jGj47dG

## Academic Experience

**Enterprises Information Laboratory**, Beijing University of Posts and Telecommunications.

*Team leader* **April, 2009 - May, 2011**

Studying on developing *Feature-Opinion Recommender System*[21] funded by the National Innovative Experimental Program.
Advisor: Prof. Guoshi Wu.

**Natural Language Processing and Computational Social Science Lab**, Tsinghua University.

*Group member* **May, 2011 - December, 2011**

Studying on social tagging at Sina micro-blog platform.
Advisor: Prof. Maosong Sun[22] and Dr. Zhiyuan Liu[23].

**The Association for Computational Linguistics**.

*Member* **June, 2012 - Present**

**Center for Speech and Language Technology**, Tsinghua University.

*Ph. D. candidate* **September, 2012 - Present**

Exploring large scale knowledge extraction approaches which can be applied to build entity-based search engine or Goolge Knowledge Graph. Advisor: Prof. Thomas Fang Zheng and Prof. Qiang Zhou.

**DolphinNLP Group & Social Card Project**.

*Founder* **June, 2012 - December, 2013**

**ACM Transactions on Intelligent Systems and Technology**[24] **(IF**[25]**: 9.39)**

*Reviewer* **November, 2014**

**2014 International Doctoral Forum**[26]

*Technical Committee Track Chair*[27] **December 5-7, 2014**

**C++ Programming**, Tsinghua University.

*Teacher Assistant* of Instructor Chao Li[28] **September, 2014 - January, 2015**

**Natural Language Processing**[29], New York University.

*Teacher Assistant* of Instructor Adam Meyers[30] **September, 2015 - December, 2015**

Special talks of "Statistical NLP: A Machine Learning Perspective"[31].

**The Eighth International Conference on Creative Content Technologies (CONTENT**

[21] http://pan.baidu.com/s/1jGj47dG
[22] http://scholar.google.com/citations?user=zIgT0HMAAAAJ&hl=en
[23] http://scholar.google.com/citations?user=dT0v5u0AAAAJ
[24] http://tist.acm.org/editors.html
[25] http://www.spinellis.gr/blog/20140808/
[26] http://www.nwpu-aslp.org/phdforum/2014/index.html
[27] http://www.nwpu-aslp.org/phdforum/2014/content/orgcommittee.html
[28] http://dbgroup.cs.tsinghua.edu.cn/lichao/
[29] http://cs.nyu.edu/courses/fall15/CSCI-UA.0480-006/
[30] http://nlp.cs.nyu.edu/people/meyers.html
[31] http://1drv.ms/1SaijfO

**2016)**[32]

*Technical Program Committee Member*[33] **March 20-24, 2016**

## Professional Experience

**Hulu Inc.**, Beijing, P.R. China.

*Occupation:* Research intern **May, 2013 - September, 2013**

- *Brief Intro*: Researcher and computer model developer for extracting relation instances from movie plots.
- *Programming Language*: `Python`.
- *Supervisor*: Post Dr. Tao Xiong.

**Microsoft Research Asia, Machine Learning Group**[34], Beijing, P.R. China.

*Occupation:* Research intern **April, 2014 - July, 2014**

- *Brief Intro*: Studying the link prediction in large-scale incomplete knowledge base, i.e. Freebase and WordNet, based on the low-dimensional embedding representation of entities and relationships without extra free texts. Developing parallel algorithms for large-scale knowledge embedding under the Microsoft Cosmos System[35].
- *Programming Language*: `C++, C#, Microsoft-SCOPE (Cosmos)`.
- *Supervisor*: Researcher Jianwen Zhang[36].

**Baidu Inc., Natural Language Group**, Beijing, P.R. China.

*Occupation:* Research intern **November, 2014 - January, 2015**

- *Brief Intro*: R&D for Baidu Chatbot (Smart Search) System[37]. Modeling/designing multi-round dialogue algorithms/system for chatbot.
- *Programming Language*: `C++`.
- *Supervior*: Senior Researcher Shiqi Zhao[38] and Dr. Rui Yan[39].

**Bosch Research**, Palo Alto, CA. U.S.

*Occupation:* Research intern authorized by NYU **June, 2015 - September, 2015**

- *Brief Intro*: Table detection and extraction from PDF files, such as electronic manuals and academic articles[40]. We also design the end-to-end system (PDFExtraction) for Bosch Research.
- *Programming Language*: `Java`.
- *Supervisor*: Senior Research Engineer, Doo Soon Kim[41].

**DailyCast Lab**, Beijing, P.R.China.

*Occupation:* Leading Researcher and Developer **May, 2016 - Present**

---

[32] `http://www.iaria.org/conferences2016/CONTENT16.html`

[33] `http://www.iaria.org/conferences2016/ComCONTENT16.html`

[34] `http://research.microsoft.com/en-us/groups/ml/`

[35] `http://blogs.msdn.com/b/seliot/archive/2010/11/05/cosmos-petabytes-perfectly-processed-perfunctorily.aspx`

[36] `http://research.microsoft.com/en-us/people/jiazhan/`

[37] http://baike.baidu.com/view/14785371.htm?fr=aladdin

[38] `http://ir.hit.edu.cn/~zhaosq/`

[39] `https://sites.google.com/site/ruiyan516/`

[40] We collect all free published articles (9,466) from ACL Anthology: `http://aclweb.org/anthology/`

[41] `https://sites.google.com/site/2soonk/`

- *Brief Intro*: Leading the research and development of short-video recommender systems in Dailycast[42] App.
- *Programming Language*: `Python` and `Java`.

## Honors and Awards

HUAGUANG 2nd Class Ph.D. Scholarship (1,000 CNY), 2015.

HUAWEI Ph.D. Fellowship (10,000 CNY), 2014.

RIIT Graduate Research Award (3,000 CNY), Tsinghua University, 2014.

Excellent Graduate Award in Tsinghua University, 2013.

National Scholarship for Undergraduates (8,000 CNY), 2011.

The IBM Chinese Excellent Undergraduate Scholarship (4,000 CNY), 2011.

Excellent Undergraduate Award in Beijing, China, 2011.

Excellent Academic Paper Award of 11th Creative Award in Beijing University of Posts and Telecommunications, 2011.

Great Leader Award of National Innovative Experimental Program, China, 2011.

MCM/ICM (Mathematical Contest in Modeling/ Interdisciplinary Contest in Modeling) **Meritorious Winner** (First Prize), 2011.

"Tang Jun&Sun Chunlan" Enterprise Scholarship (5,000 CNY), 2011.

## Participated Contests

**Kaggle Competitions** **April, 2015 - Present**

- *Profile:* `https://www.kaggle.com/michaelfan`
- *Rank:* 2,768th / 465,867 (Top 1%)
- *Awards:* Top25%(Bronze) * 2
- *Tools:* I use *NLTK*[43], *Scikit-learn*[44], *Pandas*[45], Xgboost[46], TensorFlow[47] and *Spark*[48] to process the data, build new models and make predictions.

## Favorite (Read) Books

Yoshua Bengio, Ian Goodfellow and Aaron Courville; *Deep Learning*[49], MIT Press book in preparation. I also contribute feedback on several chapters[50].

Toby Segaran; *Programming Collective Intelligence: Building Smart Web 2.0 Applications*; O'Reilly, August 2007, first edition.

Howard B. Bandy; *Quantitative Technical Analysis: An Integrated Approach to Trading System Development and Trading Management*; Blue Owl Press, January 2015, first edition.

Steven Bird, Ewan Klein and Edward Loper; *Natural Language Processing with Python*; O'Reilly, June 2009, first edition.

Rishi K. Narang; *Inside the Black Box: A Simple Guide to Quantitative and High-Frequency Trading*, John Wiley & Sons, Inc, 2013, second edition.

---

[42] `https://play.google.com/store/apps/details?id=com.jiandaola.dailycast`
[43] Natural Language Toolkit: `http://www.nltk.org/`
[44] Machine learning library in Python:`http://scikit-learn.org/stable/index.html`
[45] Python data analysis library: `http://pandas.pydata.org/`
[46] Extreme Gridient Boosting: `https://xgboost.readthedocs.org/en/latest/`
[47] Google Deep Learning Framework: `http://www.tensorflow.org/`
[48] Lightning-fast cluster computing: `http://spark.apache.org/`
[49] `http://www.iro.umontreal.ca/~bengioy/dlbook/`
[50] `http://goodfeli.github.io/dlbook/contents/acknowledgements.html`

Kevin P. Murphy; *Machine Learning: A Probabilistic Perspective*[51], MIT Press, 2012, first edition.

Christopher M. Bishop; *Pattern Recognition and Machine Learning*[52], Springer, 2007, first edition.

Yoshua Bengio; *Learning Deep Architectures for AI*[53], Foundations and Trends in Machine Learning, Vol.2 No.1 (2009) 1-127.

Wes McKinney; *Python for Data Analysis*; O'Reilly, October 2012, first edition.

Gavin Hackeling; *Mastering Machine Learning with Scikit-learn*[54]; Packt Publishing, October 2014, first edition.

Femi Anthony; *Mastering pandas*[55]; Packt Publishing, June 2015, first edition.

Luis P. Coelho, Willi Richert; *Building Machine Learning Systems with Python*[56]; Packt Publishing, March 2015, second edition.

Nick Pentreath; *Machine Learning with Spark*[57]; Packt Publishing, February 2015, first edition.

HOBBIES

Swimming, playing table tennis and the piano.

[51] http://www.amazon.com/gp/product/B00AF1AYTQ/
[52] http://www.amazon.com/gp/product/0387310738/
[53] http://www.iro.umontreal.ca/~bengioy/papers/ftml_book.pdf
[54] https://www.packtpub.com/packtlib/book/Big-Data-and-Business-Intelligence/9781783988365
[55] https://www.packtpub.com/packtlib/book/Big-Data-and-Business-Intelligence/9781783981960
[56] https://www.packtpub.com/packtlib/book/Big-Data-and-Business-Intelligence/9781784392772
[57] https://www.packtpub.com/packtlib/book/Big-Data-and-Business-Intelligence/9781783288519